White Paper Report

Report ID: 98878

Application Number: HD5097810

Project Director: Robert Morrissey (r-morrissey@uchicago.edu)

Institution: University of Chicago

Reporting Period: 6/1/2010-8/31/2011

Report Due: 11/30/2011

Date Submitted: 11/28/2011

Final White Paper

Grant HD-50978-10

Le Dictionnaire vivant de la langue française

Robert Morrissey, Project Director

University of Chicago, The ARTFL Project

11/30/2011

A. Project Activities

The following list is a comprehensive account of the major activities carried out during the grant period for the *Dictionnaire vivant de la langue française* (DVLF):

- We acquired gigabytes of text data from francophone websites around the world, which we then cleaned and loaded into databases for use in the DVLF. This data included website text, the *Corpatext* corpus[1], the *TLFi*[2], and the *Dictionnaire de synonymes*[3].

- We researched possible Web frameworks from which to build the DVLF. After surveying a few options, we decided on the Pylons framework.

- We built mockups of a variety of user interface designs and, once we settled on a scheme, built the pages using the Mako templating engine and Blueprint CSS framework[4].

- We then incorporated our definition and sentence databases with the site interface, using JavaScript and Python. We also added a "word wheel" on the left-hand side of the page to allow for dictionary "browsing".

- At this point we had a working site that anyone with the proper link could query. Evaluations at this stage revealed that we were displaying too much data, though, and that our results pages were overwhelmingly crammed with text. Consequently, we developed various methods allowing users to control how much data they saw from each dictionary, added a visualization, and limited the number of example sentences we displayed.

- Next, we developed systems to facilitate a certain "community lexicography", including a definition submission mechanism, a new database for user submissions, and an AJAX voting tool for the ranking of our example sentences.

- We also carried out experiments in ranking our example sentences automatically, basing sentence scores on the position of a given headword in a sentence, as well as calculating a sentence's similarity to its corresponding dictionary definitions.

- Finally, we prepared a conference poster, a presentation, and two very different articles about the DVLF. These efforts are detailed below and in the appendix to this report.

We modified our work plan only slightly from our original proposal. For example, we decided to use the *Corpatext* corpus for some of our example sentences instead of FRANTEXT

---

[1] http://www.lexique.org/public/lisezmoi.corpatext.htm
[2] http://atilf.atilf.fr/tlf.htm
[3] http://www.crisco.unicaen.fr/cgi-bin/cherches.cgi
[4] http://blueprintcss.org/

in order to avoid potential usage rights problems. We also abandoned work on incorporating ARTFL's old French verb conjugator into the DVLF because we were able to implement a nice widget from wordreference.com that included verb conjugations along with a great deal of information about a given word's use in idiomatic constructions.

In the course of our work on the DVLF, we have also made a number of efforts to publicize the site. These efforts include, among other things, conference presentations, articles, and a regularly updated Twitter feed. The following points represent a complete list of our publicization work:

- DVLF-branded items
  - We designed a pen with the DVLF logo and website address. This pen has been distributed at conferences, among students and faculty in the French department at the University of Chicago, and at the University of Chicago Center in Paris. A photograph is available in the appendix.
  - We also designed a t-shirt with the DVLF logo[5]. A number of ARTFL team members have purchased the shirt and worn it to conferences.
- Wikipedia article
  - We wrote a short article about the DVLF for Wikipedia[6].
- Poster at 2010 Chicago Colloquium on Digital Humanities and Computer Science
  - We presented a live demo and poster for the DVLF at this annual colloquium, hosted in 2010 by Northwestern University[7]. A link to the poster design is available in the appendix.
- MLA roundtable discussion
  - Glenn Roe presented some slides about the DVLF as part of a roundtable discussion at the 2011 MLA entitled *The Dictionary in Print and the Cloud*[8]. A link to Glenn's slides is available in the appendix.
- *Dictionaries* article
  - Tim Allen, Robert Morrissey, and Glenn Roe co-authored an article for the annual *Dictionaries: Journal of the Dictionary Society of North America*.
- *National Bulletin* article
  - Tim Allen was invited to write an article about the DVLF for a column in the quarterly *National Bulletin* of the American Association of Teachers of French

---

[5] http://bit.ly/bcxLQQ
[6] http://en.wikipedia.org/wiki/Dictionnaire_Vivant_de_la_Langue_Française
[7] http://chicagocolloquium.org/2010-dhcs-accepted-posters/
[8] http://www.mla.org/conv_listings_detail&prog_id=S056&year=2011

(AATF). The column is sponsored by the AATF Technology Commission[9] and features accounts of ways in which French teachers have been able to incorporate new technologies in classroom instruction. In this article, Tim, a former high school French teacher, suggests a number of uses for the DVLF in different classroom contexts. A link to the article text is provided in the appendix.

B. Accomplishments

The work plan that we developed for our original grant proposal largely proved to be realistic, and the following description of project activities aligns closely with our initial task plan. After acquiring large amounts of data from the web—including a wide range of text for our example sentence corpus, entries from the TLFi, and thesaurus data from the CRISCO project—we undertook a major cleaning and processing effort to produce databases that would work together under the user interface (UI) that we were building.

The development of the UI was, as expected, a major undertaking for a small group without extensive experience in modern website design. After much deliberation, we decided to build the site using the open source Pylons Web application framework and Mako template library. Our web pages consist of a combination of traditional HTML/CSS and JavaScript/AJAX. (We will discuss these technical choices further in the *Evaluation* section below.) Once these technical decisions had been made, we spent a great deal of time developing a satisfactory, comprehensible system for the display of all of the data that we possessed. Perhaps it goes without saying that we also needed to set up a digital home for the DVLF, first on one of our servers and then on a virtual machine hosted by the University of Chicago library.

It was also at this time that we began to consider methods for ranking our example sentences. We initially chose to include up to ten sentences from each of our two corpora in which the given query word was located among the first five words of the sentence. (The notion that this word might be the *theme* of the sentence comes from functional linguistics.) At a later stage we added in a mechanism that would allow our users to vote each sentence up or down and so provide a ranking for each example sentence. Depending on the number of available sentences for each word, the voting tool gives users the ability to remove a bad example sentence from display or promote a good example sentence to the top of the list. Additionally, we recently added yet another criterion to our example sentence scoring: similarity between sentences and definitions. This system assigns a score to each sentence based on the number of n-grams that overlap between a sentence and all of the definitions for a corresponding

---

[9] http://aatftech.blogspot.com/

headword. In this model, sentences and definitions are chopped up into 1-, 2-, 3-, 4-, and 5-grams, and then compared one to another. Scores are weighted so that longer matches receive more points than shorter matches. (A matching 5-gram, for example, is worth more than a matching 2-gram.) Once the analysis is complete, scores are tallied for each sentence, normalized for sentence length, and added to our example sentence database where they combine with any user-provided score to provide a final score that will indicate the placement of the sentence on the site.

In addition to this voting tool, we also encourage users to interact with the DVLF through our contributions page. This page is outlined in greater detail below, but in short, we provide DVLF users a means to contribute their own material to the site, whether it be definitions for existing words or words that do not yet have entries.

Finally, we've documented much of our work, both on internal ARTFL blogs, and in the article we wrote for the journal *Dictionaries*. Additional documentation, as well as the source code for the DVLF, is publicly available on github[10].

We do not believe that any major project goals were left unachieved in our work on the DVLF. Some aspects of the site appear differently from their original conceptions in our grant proposal, but we feel that these differences are the natural outcome of our fairly experimental work in new territory. For example, in our proposal we spoke of generating collocation clouds for each entry in the DVLF. Once we began to develop the DVLF's user interface, however, we came to the conclusion that the addition of a word cloud large enough to be useful (that is, legible) would cause our entry pages to appear rather chaotic. Word pages that contained entries from all of our available dictionaries, as well as a full complement of example sentences and synonyms, were already tending toward the overwhelming, so we decided instead to implement a cleaner, simpler time-series graph. This graph indicates word frequency over time as demonstrated in our 2,900-work, 168,000,000-word FRANTEXT corpus. A screen shot of one of these graphs is provided in the appendix to this report.

Another aspect of the DVLF that we decided to change significantly in the course of development was the user-management system. We had originally conceived of a system that would allow users to create accounts on the site so that they might be able to contribute to the site, track their input, tag words, and perhaps create various lists of words for their own reference and to share with others. As the site came together over the past year, however, we opted to streamline the process and remove nearly all of the barriers to user input. Instead of

---

[10] https://github.com/ARTFL/dodgr

forcing users to create accounts on the site, we allow users to contribute definitions, along with usage examples, via a simple web form[11] that includes a CAPTCHA for authentication purposes. Our hope is that DVLF users will appreciate the simplicity of this model and consequently feel more inclined to add content to the site when the mood strikes them. We will, of course, reevaluate this user input system if we find that our user community desires a more sophisticated means of tracking their contributions.

C. Audiences

The DVLF has seen a fairly steady increase in visitors since we began collecting statistics for the site through Google analytics in February of 2011. Beginning with approximately 20 visitors per day in early February, the DVLF now sees more than 400 visitors every day. As of November 7th, 2011, we have had 43,192 visitors who have viewed 86,350 pages on the site. Unsurprisingly, these visitors come overwhelmingly from francophone countries: France, Canada, Belgium, Switzerland, and Morocco all figure in the top ten traffic-producing countries. After France, however, it is interesting to note that our second and eighth-highest concentration of visitors (3,870) come from the United States and the United Kingdom, respectively. This figure would seem to indicate that the DVLF is being used not only by native French speakers, but also by Francophiles and French learners.

D. Evaluation

One of ARTFL's senior staff members, Glenn Roe, performed an evaluation of the project approximately halfway through its development as he was preparing slides for his presentation at the 2011 MLA meeting discussed above. Glenn found, at that point, that we were providing much of the content we had promised to deliver, but that we had too much data on display and that we needed some kind of visualization to break up all of the text. Glenn's suggestions led to the implementation of collapsible definition listings for each dictionary in the DVLF, as well as automatic truncation of long definitions. His evaluation also led to our work on the time-series graphs visible in the sidebar on the right-hand side of the results page.

We believe that the primary strength of the DVLF is its community-oriented focus. We were able to build a system that opens the French dictionary to its user community in a novel fashion that is entirely open-access. We feel that the site is easy to navigate, despite the massive amounts of data we can display to users, and that our content is trustworthy and authoritative. The project has been discovered by thousands of users across the world, and we

---

[11] http://dvlf.uchicago.edu/soumissions/

look forward to welcoming new users in the months and years to come. We also brought the ARTFL Project into new technological territory and learned a great deal about website design, branding, and lexicography.

Development of the DVLF was not, however, without its difficulties. As mentioned above, the team of developers working on the DVLF did not have a great deal of experience building dynamic, modern websites completely from scratch. We spent a great deal of time investigating Web frameworks to use for the DVLF and, when the choice came down to either Pylons[12] or Django[13], we chose Pylons because one of our developers had the ability to participate in a nearby Pylons working group. (In other words, there was no particularly compelling reason to choose one framework over the other.) Pylons had a rather steep learning curve, though, particularly for a team new to the Python programming language, and this slowed development of the DVLF initially.

Perhaps most significantly, we encountered problems with the integration of MySQL operations into the Paster server that comes packaged with Pylons. These issues caused the server to crash unexpectedly and without notification, and the problems remain difficult to explain. We eventually moved our databases to PostgreSQL, which seems to have solved our stability problems, but this transition once again slowed down development, as no one on the team had prior experience with PostgreSQL. We are not aware of any evidence that would have alerted us to these difficulties before we began to work with Pylons, and it is not clear that use of a different framework would have alleviated the problems we experienced. Nevertheless, we had hoped to have a better experience using a well-vetted, publicly available Web framework. Alternatively, we might have considered designing our own framework, though this option likely would have kept us from putting together a working site as quickly as we did.

Additional areas that we have identified as current weaknesses in the DVLF are its lack of a fully integrated source of modern French definitions, as well as a paucity of user input. To counter the lack of modern definitions, we have been cleaning up a version of the French Wiktionary that we plan to add to the site in the coming months. Moreover, we expect to see a significant increase in user input when our articles in *Dictionaries* and the *National Bulletin* are published at the end of November. These users likely represent what might be considered "power users", that is, lexicographers, dictionary enthusiasts, academic professionals, and French teachers.

---

[12] http://www.pylonsproject.org/
[13] https://www.djangoproject.com/

All of the feedback that we have received from the DVLF community has been positive or offered constructive criticism of the site. User feedback has led to, among other things, the addition of a morphology package so that the DVLF can offer spelling suggestions, as well as a more detailed user submission page. The DVLF's "mot du jour" on Twitter is regularly retweeted, and we have received a number of very kind messages in response to our tweets. The DVLF twitter account[14] currently has more than 500 followers. We were also very pleased to discover a short presentation of the DVLF on professor Jacques François's website *Interlingua*[15]*,* where he points out a number of the DVLF's distinguishing features.

We are also aware of a handful of external dictionary aggregator websites that have included the DVLF among the lexicographical resources they offer their users. One popular site, Lexilogos, links to the DVLF from its primary French dictionary page and acknowledges our display of synonyms and unique linguistic corpus[16]. Another site, Lexicool.com, includes the DVLF right along offerings from established dictionary publishers Larousse and Collins[17].

E. Continuation of the Project

As a public, "updateable" resource, the DVLF by its nature demands supervision. We plan to monitor the site continuously in order to ensure that our user-submitted definition feature is not taken advantage of by opportunistic spammers or other users who do not respect the DVLF's function as a lexicographical resource. Indeed, the quality of the DVLF's data is critical to the integrity of the ARTFL Project. We have been working to produce a customized version of the DVLF that would interface directly with our flagship databases as a JavaScript-powered definition-lookup tool. This service would replace the aging tool we have in place currently and would provide ARTFL users with a time-contextualized lookup function on top of all that the DVLF currently offers. We also hope that the proximity of example sentences in ARTFL's databases to the DVLF might lead users to add more usage examples to the site.

We also have plans to continue adding to the DVLF's wealth of data. A version of the French Wiktionary, referenced above, is nearly ready to be folded into the DVLF's dictionary offerings. This addition is particularly significant, as it represents the first fully modern dictionary available through the DVLF. (The TLFi might generally be described as a "modern" dictionary, but it is in fact decades old.) We have also gathered new data to add to our usage example corpus. This new data will expand our corpus's coverage of French as it is employed around

---

[14] https://twitter.com/#!/LeDVLF
[15] http://www.interlingua.fr/accueil/actualites/
[16] http://www.lexilogos.com/francais_langue_dictionnaires.htm
[17] http://www.lexicool.com/dictionnaire-ressources-francais.asp

the world in a variety of domains.

Finally, we recently moved the DVLF to its long-term home on a virtual machine at the University of Chicago library. This move provides the DVLF a team of expert system administrators with an unimpeachable record of reliability. We will also have a "hot-swappable" backup of the DVLF ready to deploy at any time in the unlikely event of a system failure.

F. Long Term Impact

The ARTFL Project fully expects to maintain the DVLF as one of its core French-language resources for many years to come. We've made the DVLF the default French dictionary source for our old, yet extremely popular, French-English dictionary tool[18], and, as discussed above, we have been working to connect a customized version of the DVLF to ARTFL databases.

We are also proud to mention here the development of a DVLF-inspired dictionary site for Greek and Latin, entitled *Logeion*[19]. The *Logeion* project was conceived by University of Chicago Associate Professor of Greek, and long-time ARTFL collaborator, Helma Dik, who had been looking for a way to integrate various types of corpus data with a unified, streamlined lookup tool for the plethora of Greek and Latin dictionaries she had in her possession. Using the DVLF as both a conceptual model and a technical framework, Professor Dik was able to secure funding for *Logeion* from the Women's Board of the University of Chicago. Members of the ARTFL team who worked on the DVLF provided consultation to Professor Dik's developers, who were then able to modify the DVLF's templates and lookup functions to meet their needs.

G. Grant Products

The primary product of this grant is clearly the dictionary website itself, http://dvlf.uchicago.edu/. Also, as discussed above, our work on the DVLF led to conference presentations, a general-interest article, and a scholarly article for lexicographers. (All of these products are represented in the appendix to this report.)

As awareness of the DVLF grows among French speakers, teachers, and students, however, our hope is that the final product of this grant will not simply be a website, but rather the emergence of the larger community that grows around it. Through continued maintenance and the addition of new resources, we expect that the DVLF's audience will continue to develop and that its community of regular users will expand. The data that these users provide about

---

[18] http://machaut.uchicago.edu/?resource=frengdict
[19] http://logeion.uchicago.edu/index.html

the French language, as well as about their own dictionary usage, will undoubtedly lead to site improvements and new research questions around the notion of community lexicography. The site is designed to evolve according to its users' needs, and we look forward to watching the way in which that evolution takes place.

H. Appendix

PAPERS/PRESENTATIONS

- [Seeing the dictionary again for the very first time: The *Dictionnaire vivant de la langue française*](). Article written by project manager Tim Allen for the National Bulletin of the American Association of Teachers of French. Published November 2011.
- Re-imagining French Lexicography: The Dictionnaire vivant de la langue française. Article written by Tim Allen, Robert Morrissey, and Glenn Roe for *Dictionaries: Journal of the Dictionary Society of North America*, volume 32 (2011).
- [Slides]() from Glenn Roe's presentation of the DVLF as part of the roundtable panel "The Dictionary in Print and in the Cloud", organized by the Dictionary Society of North America at the 2011 Modern Language Association conference.
- [DVLF Poster]() presented at 2010 Chicago Colloquium on Digital Humanities and Computer Science, hosted by Northwestern University
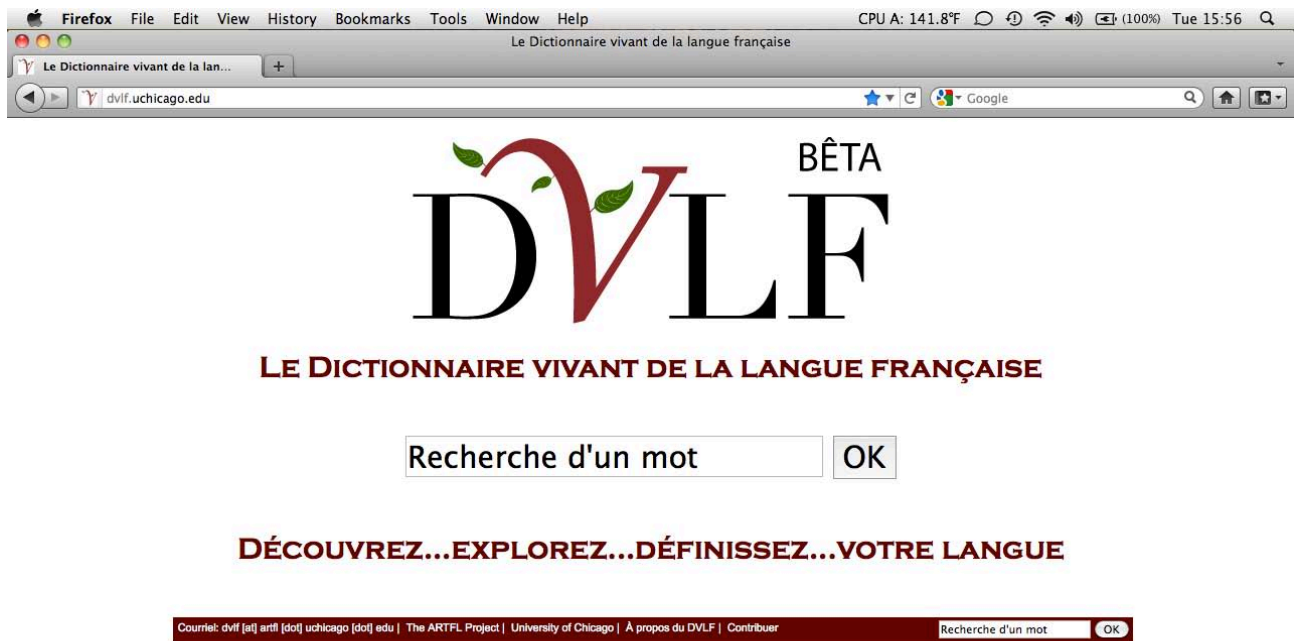
BRANDING

DVLF pen:

DVLF t-shirt:

SCREEN SHOTS

Home page:

Results page:

ARTFL :: Dictionnaire vivant de la langue française :: vivace

ARTFL :: Dictionnaire vivant de l...   +

dvlf.uchicago.edu/mot/vivace   Google

DVLF BÊTA   **DICTIONNAIRE VIVANT DE LA LANGUE FRANÇAISE**

Recherche d'un mot   OK

Index alphabétique

vitriolisation
vitrosite
vittadine
vitteau
vituaille
vitulaire
vitulin
vituperable
vitupérable
vitupératif
vitupéré
vitupere
vitupère
vitupérer
vituperer
vitupereur
vitupéreux
vitupération
vitupérer
vivable
vivace
vivacement
vivacité
vivacité
vivandier
vivaneau
vivant
vivarium
vivat
vive

vivace

12 définitions trouvées dans 7 dictionnaires.

▶ LE TRÉSOR DE LA LANGUE FRANÇAISE INFORMATISÉ   1

▽ DICTIONNAIRE DE L'ACADÉMIE FRANÇAISE 8E ÉDITION (1932-1935)   1

- adj. des deux genres. Qui a une grande vitalité. *Le corbeau, le cerf sont des animaux vivaces.* Il se dit figurément de Ce qui est de longue durée ou difficile à détruire. *Les préjugés sont vivaces.* En termes de Botanique, il se dit des Plantes qui repoussent [...]

▶ ÉMILE LITTRÉ: DICTIONNAIRE DE LA LANGUE FRANÇAISE (1872-1877)   6

▶ DICTIONNAIRE DE L'ACADÉMIE FRANÇAISE 6E ÉDITION (1835)   1

▶ DICTIONNAIRE DE L'ACADÉMIE FRANÇAISE 5E ÉDITION (1798)   1

▶ FÉRAUD: DICTIONAIRE CRITIQUE DE LA LANGUE FRANÇAISE (1787-1788)   1

▶ DICTIONNAIRE DE L'ACADÉMIE FRANÇAISE 4E ÉDITION (1762)   1

Vous aimez une phrase ? Faites valoir votre opinion, votez pour elle !
8 phrases trouvées dans 2 corpus.

▽ CORPATEXT   2

⬆ 2 ⬇ Tandis que le jeune et **vivace** amour à l' infini se prismatise de nuances comme un beau ciel, une tranquille rivière, une florale prairie, les stériles fatigues de la chair n' ont qu' une note toujours la même.

⬆ -1 ⬇ Cette ronce était de l'espèce **vivace** dite mûrier de renard.

▽ WWW   6

Prononciation
vivas

Synonymes
solide, tenace, durable, vif, résistant, vigoureux, vivant, indestructible, ineffaçable, robuste, coriace, dur, enraciné, étincelant, généreux, impérissable, invétéré, persistant, pluriannuel, présent, rustique, violent, vivant

Antonymes
annuel, caduc, flétri

Usage à travers le temps
Occurrences du mot "vivace" pour un million de mots

Activer WordReference

Close-up of sidebar widgets:

Prononciation

vivas

Synonymes

solide, tenace, durable, vif, résistant, vigoureux, vivant, indestructible, ineffaçable, robuste, coriace, dur, enraciné, étincelant, généreux, impérissable, invétéré, persistant, pluriannuel, présent, rustique, violent, vivant

Antonymes

annuel, caduc, flétri

Usage à travers le temps

Occurrences du mot "vivace" pour un million de mots

WordReference.com
Dictionaries

French–English   Go

Concise Oxford-Hachette French Dictionary © 2009 Oxford University Press:

vivace /vivas/
adjective
1. *Bot* **(plante)** ~ perennial;
2. *(durable)* enduring.

'vivace' found in these entries:
French:
  plante - vivant
English:
  perennial - rye - unspent

vivace: WordReference English-French Dictionary © 2011

vivace *adj* (vibrant, vigoureux) *vibrant, vigoureux*		vivacious
vivace *nf* (plante) *plant*		hardy perennial
		Report an error

vivace: WordReference English-French Dictionary © 2011

Compound Forms/Formes composées

ivraie vivace *nf* (plante, botanique)	ryegrass
ivraie vivace	perfect binding
ivraie vivace	rye grass

14

Close-up of example-sentence listing:

Vous aimez une phrase ? Faites valoir votre opinion, votez pour elle !

8 phrases trouvées dans 2 corpus.

▽ **CorpaText**   2

⬆ **2** ⬇ Tandis que le jeune et **vivace** amour à l' infini se prismatise de nuances comme un beau ciel, une tranquille rivière, une florale prairie, les stériles fatigues de la chair n' ont qu' une note toujours la même.

⬆ **-1** ⬇ Cette ronce était de l'espèce **vivace** dite mûrier de renard.

▽ **WWW**   6

⬆ **1** ⬇ Une tradition américaine particulièrement **vivace** à Los Angeles depuis la fuite d'OJ Simpson sous les caméras héliportées en 1994. (www.liberation.fr)

⬆ **0** ⬇ Album concept de chair **vivace** et de larmes retenues, The Boat percute de ses atmosphères forcément chaloupées l'univers maladif de Matt Elliott. (octopus)

⬆ **0** ⬇ L'histoire reste **vivace** dans des villes et de villages tels qu'Alacala de Henares, lieu de naissance de Cervantes, Sigüenza, qui conserve une cathédrale romane, et Chinchon où les Madrilènes viennent acheter le vin et l'anis locaux sous les galeries en bois d'une grand-place médiévale. (fr.canoe.ca)

⬆ **0** ⬇ L'insigne toujours **vivace** du Parti Communiste Chinois est sombre et rouillé. (www.liberation.fr)

⬆ **0** ⬇ Le souvenir reste **vivace** de la fuite devant les armées d'un Etat d'Israël en formation. (www.liberation.fr)

⬆ **0** ⬇ Le souvenir demeure **vivace** aux Etats-Unis de l'attaque suicide à la bombe contre la caserne des Marines à Beyrouth, qui avait tué 241 militaires américains en 1983, des attentats à répétition contre l'ambassade US à la fin des années 80 et des prises d'otages étrangers dont les auteurs n'ont jamais été arrêtés. (www.liberation.fr)